An automatic image-based system for detecting wild and stocked fish

Espen Myrum*, Simen Andre Nørstebø*, Sony George*, Marius Pedersen*, and Jon Museth+
* Department of Computer Science, NTNU
+ Norwegian Institute for Nature Research

Abstract

Fish stocking is the method of raising fish in a hatchery and releasing them into a river or lake to sustain or increase an existing population or to create a population. This has been practised in many countries, including Norway. Before the fish are released, the adipose fin is commonly removed in order to identify that it is a stocked fish. Cameras have been mounted in several Norwegian rivers in order to monitor fish populations. Classification of fish from these cameras is today a manual task carried out by people. In this paper we propose an automatic classification method to separate wild fish from stocked fish using machine learning. Experiments on an image set of trouts (Salmo Trutta) show a very high accuracy of the proposed method.

1 Introduction

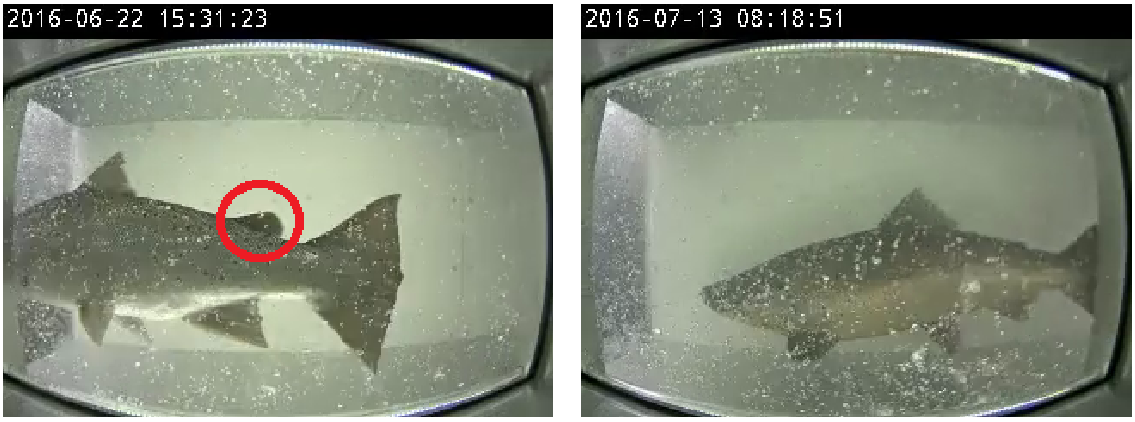
Many rivers and lakes are not able to sustain a population of fish or one would like to establish a fish population in a river or lake where there is not a fish population. Fish stocking has been a method to do this, where fish are raised in a hatchery and then released into the wild when they reach a certain size. Stocked fish normally have their adipose fin, a small fleshy fin found between the dorsal fin and the caudal fin, removed prior to being released in the wild, being the best way to identify stocked fish from wild fish (see Figure 1).

There is a need to monitor fish populations to contribute to the knowledge base for more sustainable management of rivers and lakes. A common way to monitor fish populations is by using an underwater camera, which is installed in a specially designed setup. In rivers this is commonly installed in a fish ladder, which is a structure on or around natural or man-made barriers (for example dams or waterfalls). This allows to monitor all fish migrating through the fish ladder. When a fish passes a sensor placed in the fish ladder, the camera starts to record a video. These videos can be used to determine the fish species of the fish passing the camera, but also to determine whether the fish is wild or stocked. Today, this work is commonly done manually by a human, which is time and resource demanding. Classification of whether a fish is raised in the wild or a hatchery is

This paper was presented at the NIK-2019 conference; see http://www.nik.no/.

important information that can benefit population count, quality control of fish and also monitoring the ecosystem. In this work we focus on automatic classification of fish in underwater images to decide whether it is wild or stocked. We will focus on the trout species (Salmo trutta), in images captured in freshwater.

Living fish classification is a challenging problem since the fish move freely, light conditions can vary significantly, different visibility in the water, and objects that are not desired can occur (non-fish objects, for example leaves, branches, etc.). Good quality video is required, especially when making a decision on whether or not a fish is wild or stocked, as it is only the adipose fin that can provide information on this. Figure 1 shows example images from an underwater camera, where Figure 1a shows a wild trout with the adipose fin intact while Figure 1b shows a stocked fish with the adipose fin removed. As these fish in the example images are larger fish, it is easier to see the adipose fin. Smaller fish makes it more difficult, and also requiring better video quality.



(a) Wild trout (b) Stocked trout

Figure 1: Example image of a wild trout and stocked trout. The adipose fin is marked in a red circle on the wild trout on the left, while the trout on the right does not have an adipose fin and is therefore a stocked trout.

The goal of this paper is to classify trout as wild or stocked fish from underwater natural images, and we will do this by using a deep neural network and transfer-learning.

This paper is organized as: first we present relevant background in Section 2, then we continue with materials and methods in Section 3, further results and discussion are presented in Section 4, at last we conclude and present future work in Section 5.

2 Relevant background

To the best of our knowledge automatic classification of wild and stocked fish has not been published before. However, there exist work on classification of fish species and the use of machine vision systems in aquaculture [1].

In underwater environments, many approaches for classification have been presented during the last decade. Ogunlana et al. [2] used a Support Vector Machine (SVM) based technique to solve binary fish classification based on 6 shape features with a very small training data. The results showed an accuracy of almost 80%. Chuang et al. [3] proposed a fully unsupervised clustering approach for multiclass classification based on SVM, binary hierarchy and partial classification. Their results also show a high accuracy on fish image dataset.

Boom et al. [4] introduced an underwater camera surveillance system for monitoring fish. They used a hierarchical classifier based on color, contour and texture features for classification of fish. They stated to be able to recognize 15 different fish species, and showed an average recall of about 83%. Hossain et al. [5] classified different fish species using pyramid histogram of visual words features with an SVM classifier. They evaluated their method on low quality and high quality images showing an accuracy of 40.1% and 91.7%, respectively.

Villon et al. [6] introduced two methods for fish classification; the first based on a traditional two-step approach with extraction of histogram of gradient features and an SVM classifier, and the second method based on deep learning with the GoogLeNet architecture. The results from their experimentation showed the first method to give an F-measure of 0.49 and for second method an F-measure of 0.64.

Pengying et al. [7] proposed a Convolutional Neural Network (CNN) based classification method for trout and grayling. Their approach was based on the pre-trained Alexnet with stochastic gradient descent with momentum. Their experiments showed a very high accuracy above 99%, their method could also classify incomplete fish images with an accuracy of 98%. They also evaluated if Contrast-limited adaptive histogram equalization (CLAHE) would improve the classification accuracy, but it was shown to be decreased by pre-processing.

Pre-processing has also been introduced for underwater classification. Rizzini et al. [8] used CLAHE to compensate light attenuation and remove artefacts in the images before applying a multi-feature object detection algorithm. They state to have satisfactory precision and recall on three different datasets.

Zhao et al. [9] proposed an image-based solution for recognition of individual trouts in the wild. They used the spot pattern from a region from the head for recognition, as this region has shown to be enough for individual fish recognition [10, 11]. Their approach was based on a codebook, where Speeded Up Robust Features (SURF) were used to generate the codebook. Further, SVM was used to generate the final output descriptors. Their evaluation showed an accuracy of about 74% when it came to individual recognition.

Recently, CNNs and deep learning have also been applied to underwater fish classification. Rathi et al. [12] combined CNNs, deep learning and image processing to deal with background noise, image distortion, undesirable objects, occlusion and image quality. On a 23 fish species dataset they obtained 96.3% accuracy.

Machine vision has been used in aquaculture for monitoring fish. This has been driven by the advantage of being fast, cheap and noninvasive. Odone et al. [13] used image parameters to find a relationship between weight and shape, this was done through using SVM. De Verdal et al. [14] used image analysis for individual growth monitoring. Zion et al. [15] used fish area to estimate fish mass, but mentions that image quality needs to be high enough for accurate segmentation and shape analysis. Low quality images makes it difficult to separate between two fish species in their study. Bermejo [16] tested different support vector machine classifiers using a cod database for fish age classification, and showed that a combination of fish length, weight sex with morphological features gave an accuracy of about 75%.

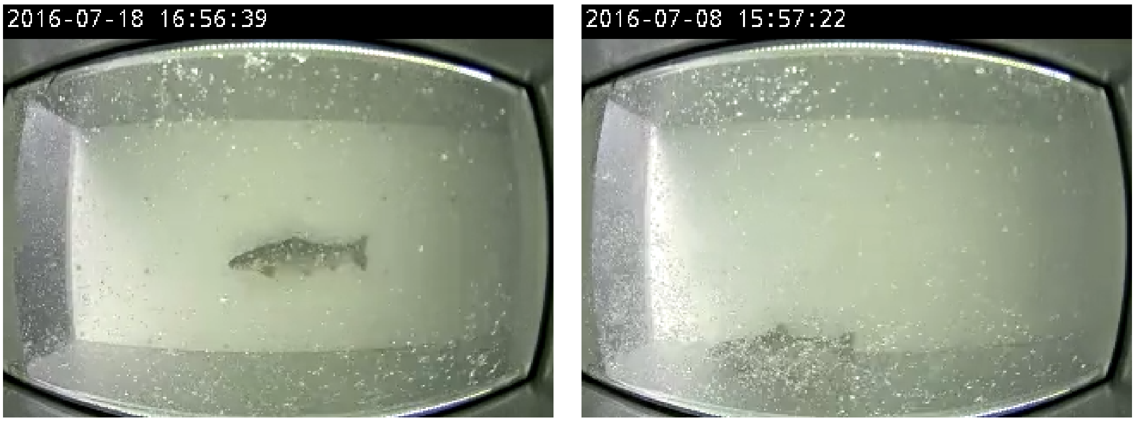
3 Materials and methods

We will first present the dataset used in this study before we introduce the proposed method for automated classification of wild and stocked trout.

Dataset

Our dataset contains 204 video clips, where 101 videos are of stocked fish and 103 videos of wild fish. Each video clip is 24 seconds with a resolution of 320×240 pixels. The quality of the videos are varying in terms of illumination level, illumination uniformity, as well as distortions as air bubbles (as shown in Figure 2 on the right side) and algae. Examples of a "good" image is shown in Figure 3a and an image that is not as sharp, with discoloration of the water and with more air bubbles is shown in Figure 3b. The videos contain fish of different sizes, from small to large trouts (Figures 3b and 2. These videos provide a diverse dataset with challenges. We expect smaller trouts to be more difficult to classify as their adipose fin is not as visible compared to larger trouts. Also the combination of air bubbles, discoloration and small fish would be difficult in terms of classification.

These videos have been converted into still images, where each frame is a still image. The format of the images is PNG.



(a) (b)

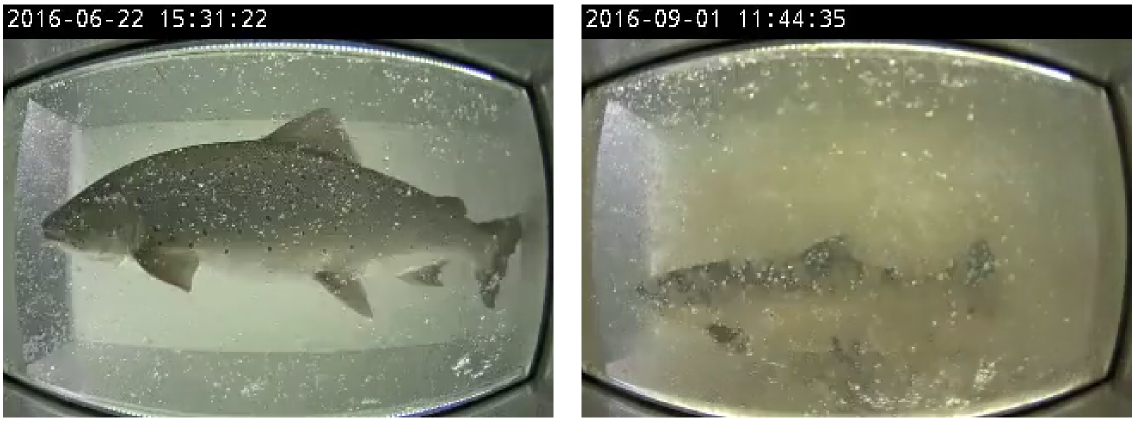
Figure 2: Example images. A smaller trout on the left without occluding air bubbles, and on the right a smaller trout with many occluding air bubbles. It is clear that air bubbles can influence the classification.

The images have been categorized into three classes by an expert serving as the ground truth; no fish (images with no fish), stocked fish and wild fish. Our dataset contains 5000 images of each class.

Data augmentation has been done through a horizontal flip of the images, doubling the dataset. A horizontal flip also makes sense as the fish can swim upstream and downstream through the fish ladder. We split the dataset in two parts for training and testing, namely 70% for training and 30% for testing.

Proposed method

Our starting point is the pretrained AlexNet convolutional neural network, which is trained on more than a million images from the ImageNet database [17]. Alexnet is eight layers deep, consisting of 5 convolutional layers and 3 fully connected layers,



(a) (b)

Figure 3: Example images. A higher quality image on the left and a lower quality image on the right. It is clear that discoloration can influence the classification.

and has been trained to classify images into 1000 object categories. Because of this Alexnet has learned many feature representations for a variety of images. Image input to the network is 227 by 227 pixels. Our approach is by using transfer learning and fine-tuning the network to the application of classifying wild and stocked fish. We refer to this approach as the pretrained Alexnet CNN hereafter.

For the pretrained Alexnet CNN we tested different optimizers, namely ADAM [18], stochastic gradient descent with momentum (SDGM), and RMSprop. We also tested if pre-processing the images with Contrast-limited adaptive histogram equalization (CLAHE) [19] would improve the classification. Our tests showed that SDGM gave the best results without CLAHE. Results shown hereafter is by using SDGM on the original images (without any pre-processing).

For the training we used a batch size of 32, 15 epochs, initial learning rate of 0.003, and we reduced the learning rate by a factor of 0.1 every 5 epochs. For the reported results we show the average results after training the network five times.

Another approach we will test is based on an Error correcting output codes (ECOC)-classifier. ECOC is an ensemble method specifically designed for multi-class classification. It uses binary classifiers to solve a multi-class problem. After transfer learning using the pretrained Alexnet CNN, it is run through the ECOC classifier. We refer to this approach as ECOC hereafter.

The implementation has been done in Matlab R2018B, using the "Deep Learning Toolbox 12.0" and the "Statistics and Machine Learning Toolbox 11.4".

4 Results and discussion

For the pretrained Alexnet CNN we obtain an average accuracy of 99.16% and for the ECOC approach we obtained an accuracy of 99.87%. Observations of the results indicate that the misclassification occurs when the trouts are small, which makes it more difficult to see the adipose fin. In some frames there are a significant amount of bubbles (distortions) which also makes classification difficult, and discoloration could also impact. These aspects are illustrated in Figure 2 and Figure 3. The findings that bubbles influence the classification is not surprising, and it is also found in the literature that noise [20, 21] (which is somewhat similar to bubbles) will decrease the performance of neural networks. Experts from NINA also confirmed that bubbles is a problem when visually classifying if a trout is wild or stocked.

We also report the result for each video in the dataset, where we have calculated the average probability for the frames in each video for a given class. We will use 80% as a threshold, meaning that an average probability of the class above 80% is considered as good. The results can be shown in Table 1. We can see that for the pretrained Alexnet CNN for the wild fish 94 videos (out of 103) have an average probability of belonging to that class to be higher than 80%, 9 videoes are below but still correctly predicted and none are misclassified. The results for ECOC for wild fish are similar. When we see the results for stocked fish, the pretrained Alexnex CNN has 67 videos at a probability above 80%, 33 videos below 80% but still predicted as stocked fish, and one misclassified video. ECOC reports better numbers with 87 videos above 80%, 14 below 80% and none misclassified. These results might indicate better results for ECOC than the pretrained Alexnet CNN.

Method	Wild fish			Stocked fish		
	higher than 80%	less than 80%	Mis-classified	higher than 80%	less than 80%	Mis-classified
Pretrained Alexnet CNN	94	9	0	67	33	1
ECOC	90	13	0	87	14	0

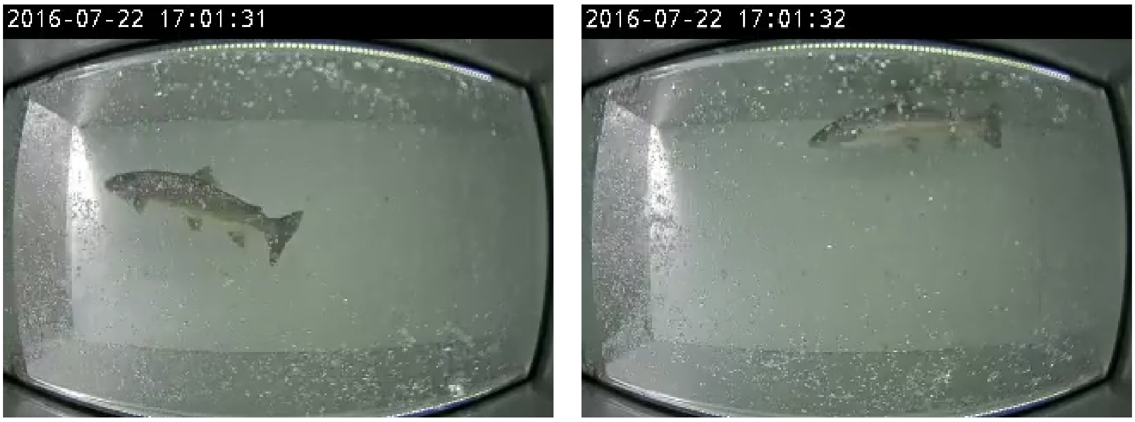
Table 1: Classification results. We can see that both methods (pretrained Alexnet CNN and ECOC) provide good results. ECOC does not have any misclassified fish.

Visual investigation shows that videos with a lot of air bubbles are challenging, but also frames with low quality (blurred and/or with low contrast). An example of a challenging frame that is blurred and low contrast is shown in Figure 4. Blur and low contrast can be caused by different factors; motion blur occurs when fish swims very fast through the fish ladder, faster water flow and disturbance in the water will also impact the sharpness, and low illumination can influence both. One observation from the dataset is that some videos contain algae and debris in the frames, although this is not changing in the period that the fish is passing through the field of view, there is a significant difference between videos. Algae can be seen in Figure 4. We have not observed that this influenced our results, but it could potentially have some impact. Algae will slowly be introduced over time in these videos, but when the camera setup is cleaned it will introduce a significant difference.

Visually there are more air bubbles in the top and bottom part of the frame, and it seems like fish that passes through the center of the frame is classified with a higher accuracy than those passing at the bottom or top. An example of this aspect is shown in Figure 5. At last, frames where the fish is closer to the lens will results in the fish covering a larger area, which could also make the detection of the adipose fin easier. A larger dataset is required to verify many of the aspects discussed here.

We also tested how JPEG compression influenced the performance, where the network was trained on PNG images and tested on JPEG images. The performance was almost identical to that for testing on PNG images. This corresponds well to the findings of Dodge and Karam [20], where they found CNNs to be very resilient towards JPEG compression, as long as the compression is not at very low quality levels. There has also been proposed work to be more robust towards JPEG

Figure 4: Example of challenging frame that is blurred and with low contrast. Algae can also be seen in the frame.



(a) (b)

Figure 5: Example of how air bubbles influence the frames. The two frames are from the same video of the same trout. When the fish is at the top of the frame occluding air bubbles can influence the classification. Less occlusion is present when the fish is closer to the center.

compression and other distortions [22], which could be a direction for future work.

5 Conclusion and future work

We have investigated the use of deep learning for automatic classification of stocked and wild trouts (Salmo Trutta). Two different approaches; one based on transfer learning using a pretrained Alexnet convolutional neural network and the other being similar to the first but with an Error-Correcting Output Codes classifier. The results for both give very good validation accuracy; 99.16% and 99.87%, respectively. However, the dataset contains only 5000 images in each class, which is a limitation. Despite the limitations, this is to be of our knowledge the first work at trying to automate the classification of wild and stocked fish.

Our approach was based on still images, the work should be extended to work on video, as well as dealing with multiple fish in the same video frame. Additional evaluation should also be done on a larger dataset with extensive statistical evaluation. Other fish than trouts should also be considered.

References

[1] Mohammadmehdi Saberioon, Asa Gholizadeh, Petr Cisar, Aliaksandr Pautsina, and Jan Urban. Application of machine vision systems in aquaculture with emphasis on fish: state-of-the-art and key issues. *Reviews in Aquaculture*, 9(4):369–387, 2017. doi: 10.1111/raq.12143. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/raq.12143`.

[2] SO Ogunlana, O Olabode, SAA Oluwadare, and GB Iwasokun. Fish classification using support vector machine. *African Journal of Computing & ICT*, 8(2):75–82, 2015.

[3] M. Chuang, J. Hwang, and K. Williams. A feature learning and object recognition framework for underwater fish images. *IEEE Transactions on Image Processing*, 25(4):1862–1872, April 2016. ISSN 1057-7149. doi: 10.1109/TIP.2016.2535342.

[4] Bas Boom, P.X. Huang, Concetto Spampinato, S. Palazzo, Jiyin He, C. Beyan, Emmanuelle Beauxis-Aussalet, Jacco van Ossenbruggen, G. Nadarajan, Yun-Heh (Jessica) Chen-Burger, Daniela Giordano, Lynda Hardman, Fang-Pang Lin, and Robert Fisher. Long-term underwater camera surveillance for monitoring and analysis of fish populations. In *Workshop on Visual observation and Analysis of Animal and Insect Behavior 2012*, pages 1–4, jan 2012.

[5] Ekram Hossain, SM Shaiful Alam, Amin Ahsan Ali, and M Ashraful Amin. Fish activity tracking and species identification in underwater video. In *2016 5th International Conference on Informatics, Electronics and Vision (ICIEV)*, pages 62–66. IEEE, 2016.

[6] Sébastien Villon, Marc Chaumont, Gérard Subsol, Sébastien Villéger, Thomas Claverie, and David Mouillot. Coral reef fish detection and recognition in underwater videos by supervised machine learning: Comparison between deep learning and hog+ svm methods. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 160–171. Springer, 2016.

[7] Thitinun Pengying, Marius Pedersen, Jon Yngve Hardeberg, and Jon Museth. Underwater fish classification of trout and grayling. In *Under review*, 2019.

[8] Dario Lodi Rizzini, Fabjan Kallasi, Fabio Oleari, and Stefano Caselli. Investigation of vision-based underwater object detection with multiple datasets. *International Journal of Advanced Robotic Systems*, 12(6):77, 2015.

[9] Lingcong Zhao, Marius Pedersen, Jon Yngve Hardeberg, and Børre Dervo. Image-based recognition of individual trouts in the wild. In *8-th European Workshop on Visual Information Processing*, Oct 2019.

[10] Carlos Garcia de Leaniz, Neil Fraser, Victor Mikheev, and Felicity Huntingford. Individual recognition of juvenile salmonids using melanophore patterns. *Journal of Fish Biology*, 45(3):417–422, 1994.

[11] Caitlin M Gifford and David W Mayhood. Natural marks for identifying individual fish in small populations of at-risk westslope cutthroat trout. In

Wild trout symposium IX: sustaining wild trout in a changing world, pages 275–81, 2013.

[12] Dhruv Rathi, Sushant Jain, and S. Indu. Underwater fish species classification using convolutional neural network and deep learning. *2017 Ninth International Conference on Advances in Pattern Recognition (ICAPR)*, pages 1–6, 2017.

[13] Francesca Odone, Emanuele Trucco, and Alessandro Verri. A trainable system for grading fish from images. *Applied Artificial Intelligence*, 15(8):735–745, 2001.

[14] Hugues De Verdal, Marc Vandeputte, Elodie Pepey, Marie-Odile Vidal, and Béatrice Chatain. Individual growth monitoring of european sea bass larvae by image analysis and microsatellite genotyping. *Aquaculture*, 434:470–475, 2014.

[15] B Zion, A Shklyar, and I Karplus. Sorting fish by computer vision. *Computers and electronics in agriculture*, 23(3):175–187, 1999.

[16] S. Bermejo. Fish age classification based on length, weight, sex and otolith morphological features. *Fisheries Research*, 84:270–274, April 2007.

[17] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[18] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[19] Karel Zuiderveld. Contrast limited adaptive histogram equalization. In *Graphics gems IV*, pages 474–485. Academic Press Professional, Inc., 1994.

[20] Samuel Dodge and Lina Karam. Understanding how image quality affects deep neural networks. In *2016 eighth international conference on quality of multimedia experience (QoMEX)*, pages 1–6. IEEE, 2016.

[21] Samuel Dodge and Lina Karam. Quality resilient deep neural networks. *arXiv preprint arXiv:1703.08119*, 2017.

[22] Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. In *International Conference on Learning Representations*, 2019. URL `https://openreview.net/forum?id=HJz6tiCqYm`.